The Software Alliance
BSA

バイアスに挑む：AIの信頼性構築に向けたBSAのフレームワーク

人工知能（AI）が飛躍的に進歩したことにより、テクノロジーによって世界が様変わりする過程において人々の期待が急速に変化しており、公平性に関する議論が進んでいます。AIは、よりよい社会を実現するための影響力となり得ますが、一方で既存の社会的バイアスが長期化（さらには悪化）し、歴史的に疎外されてきたコミュニティの人々がシステム上不利になる可能性があるとの認識が高まっています。AIが業務プロセスに搭載され、人々の生活に多大な影響を持つ現在、組織が意図しないバイアスの潜在的リスクを考慮してシステムの設計および導入を行うようにすることが非常に重要になっています。

このフレームワークは、AIが設計により確実に説明可能であるようにするためのツールであり、多種多様な組織が使用することで、システムのライフサイクルを通してバイアスによるリスクを管理することができます。このフレームワークは、膨大な研究に基づき、主要なAI開発者の経験から得られた知見を取り入れて構築されており、以下の内容が含まれています。

バイアスの潜在的リスクを特定し、軽減するための影響評価を行うプロセスについて、概要を説明します。

AIシステムのライフサイクルを通して、AIに固有のバイアスリスクが発生するおそれがあります。そのようなリスクを軽減するための、既存のベスト・プラクティス、テクニカルツール、リソースについて確認します。

AIリスク管理プログラムを効果的に実施し、支援する上で必要となる、企業ガバナンスの重要な仕組み、プロセス、保護措置を提示します。

組織はこのフレームワークをプレイブックとして使用し、リスク管理プロセスを通してAIシステムの信頼性を高め、公平性、透明性および説明責任を高めることができます。AIシステムを開発する組織や、そのようなシステムを取り入れて運用する企業はこのフレームワークを活用し、以下のような業務の基盤とすることができます。

- **内部プロセスの指針：**このフレームワークをツールとして使用することにより、内部リスク管理プロセスにおける役割、責任および期待を組織立てて定めることができます。
- **トレーニング、意識向上および教育：**このフレームワークを使用することにより、AIシステムの開発や使用に携わる従業員のための社内トレーニングや教育プログラムを構築したり、組織がAIバイアスリスクを管理するアプローチについて経営幹部を教育したりすることができます。
- **サプライチェーンの保証と説明責任：**AI開発者やAIシステムを導入している組織はこのフレームワークを基盤として使用することにより、システムのライフサイクル全体を通じてAIリスクを管理する上でのそれぞれの役割や責任について、コミュニケーションや調整を行うことができます。
- **信頼と自信：**このフレームワークは、企業が製品の特徴や、その製品がAIバイアスリスクを軽減するアプローチについて、一般の人々に情報を伝える際に役立ちます。そういった意味でこのフレームワークは、組織が倫理的なAIシステムを構築する取り組みについて一般に説明する上でも有益であるといえます。
- **インシデントへの対応：**突然のインシデントが発生した場合、このフレームワークに記載されているプロセスや参照資料が監査証跡の役割を果たし、組織が潜在的な問題を迅速に特定して修正することが容易になります。

レポート全文は、www.bsa.orgにてダウンロードすることができます。